De novo binding prediction of peptides to MHC class I

Oriol Gracia[#1*], Pep Amengual-Rigo[#2*], Victor Guallar [#&3]

[#]*Barcelona Supercomputing Center (BSC)*

[&]*ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain*

{[1]oriol.gracia, [2]jose.amengual, [3]victor.guallar}@bsc.es

**Both authors contributed equally to this work*

***Keywords*— Cancer, personalized medicine, immunotherapy, neoantigen, MHC class I**

Introduction

An emerging therapeutic application in immunology is the development of vaccines against cancer by using neoepitopes. A neoepitope is a peptide found only in the cancer state (coming from different gene expression or containing a mutation), that is able to generate an immune response. However, neoantigen discovery is hampered by the huge complexity of the immune system and the lack of experimental determinations.

The immune system is able to elicit an immune response against foreign antigens. Generation of the immune response depends on i) the degradation of proteins by the proteasome, and ii) the presentation of those peptides to the T-cells by the major histocompatibility center (MHC). Particularly in cancer, MHC class I role is to present self peptides to the CD8+ T-cells. There are described more than 10,000 MHC class I alleles[1], and it is expected that each of them is able to present around 5,000 different peptides. In contrast, a small fraction of peptides for less than 150 MHC class I alleles have been experimentally addressed. Current machine learning state-of-the art MHC binding predictors[2,3] are limited to deal only with the well characterized MHC class I alleles, hampering their application in personalized medicine. In this context we present uMHCpred, a de novo binding predictor that could give insights in the building of an universal model able to deal with the underrepresented MHC class I alleles.

Methodology

uMHCpred follows a frequency based approach that integrates the tridimensional structure of the MHC class I. More in detail, uMHCpred factorizes each residue of the peptide, building a position independent contact map. Then, the obtained contact map is used to classify each MHC by its structural similarity for each of the actorized peptide positions, archiving in this way a model that is no longer allele dependent.

In this way, uMHCpred is able to predict de novo MHC class I alleles using training data coming from those alleles that have a similar binding environment. In order to assess the performance of the de novo predictions, we compared them with a model containing all available data by computing the Receiver Operator Characteristic (ROC) and its Area Under the Curve (AUC).

Results

Our results showed that uMHCpred can predict de novo binding of MHC alleles, such as HLA-A*02:01 (**Figure 1**), reaching an accurate prediction (AUC of 0.88), with a minimal loss in comparison with the model that also uses data from the predicted allele (AUC of 0.91).

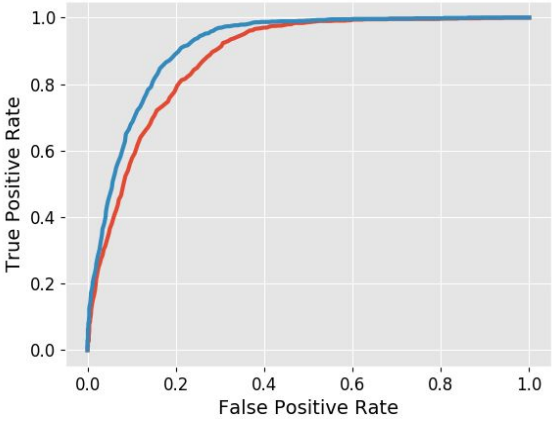


***Figure 1**. ROC curves of the de novo prediction of HLA-A*02:01 (red) and the prediction that also uses data from the same allele (blue).*

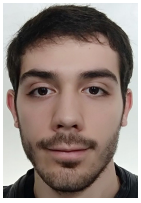
Conclusions

uMHCpred can make accurate predictions for all the MHC variants including the ones that are underrepresented in the available data. Additionally, it is able to predict which peptides could be the best theoretical binders of each allele. Furthermore, this results could open a world of possibilities towards the neoepitope finding for personalized medicine.

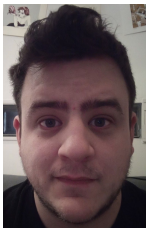
References

1. HLA Nomenclature @ hla.alleles.org. Available at: http://hla.alleles.org/nomenclature/stats.html. (Accessed: 2nd April 2019)
2. O'Donnell, T. J. *et al.* MHCflurry: Open-Source Class I MHC Binding Affinity Prediction. *Cell Syst* **7**, 129–132.e4 (2018).
3. Jurtz, V. *et al.* NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *The Journal of Immunology* 199, 3360–3368 (2017).

Biography



Oriol Gracia: BSc in Biochemistry (UB), MSc student in Bioinformatics for the Health Sciences (UPF), currently doing his Master's Thesis in the BSC supervised by Prof. Victor Guallar. His main interests are structural bioinformatics and binding energy prediction.



Pep Amengual-Rigo: BSc in Biochemistry (UIB), MSc in Bioinformatics (UAB), currently PhD student in UB under supervision of Prof. Victor Guallar at the BSC. His research lines are focusing antibody design and neoepitope prediction.